Counting substructures I: color critical graphs

Dhruv Mubayi *

April 22, 2010

Abstract

Let F be a graph which contains an edge whose deletion reduces its chromatic number. We prove tight bounds on the number of copies of F in a graph with a prescribed number of vertices and edges. Our results extend those of Simonovits [8], who proved that there is one copy of F, and of Rademacher, Erdős [1, 2] and Lovász-Simonovits [4], who proved similar counting results when F is a complete graph.

One of the simplest cases of our theorem is the following new result. There is an absolute positive constant c such that if n is sufficiently large and $1 \leq q < cn$, then every n vertex graph with $\lfloor n^2/4 \rfloor + q$ edges contains at least

$$q \left\lfloor \frac{n}{2} \right\rfloor \left(\left\lfloor \frac{n}{2} \right\rfloor - 1 \right) \left(\left\lceil \frac{n}{2} \right\rceil - 2 \right)$$

copies of a five cycle. Similar statements hold for any odd cycle and the bounds are best possible.

1 Introduction

Mantel [5] proved that a graph with n vertices and $\lfloor n^2/4 \rfloor + 1$ edges contains a triangle. Rademacher extended this by showing that there are at least $\lfloor n/2 \rfloor$ copies of a triangle. Subsequently, Erdős [1, 2] proved that if $q < cn$ for some small constant c, then $\lfloor n^2/4 \rfloor + q$ edges guarantees at least $q\lfloor n/2 \rfloor$ triangles. Later Lovász and Simonovits [4] proved that the

*Department of Mathematics, Statistics, and Computer Science, University of Illinois, Chicago, IL 60607. email: mubayi@math.uic.edu; research supported in part by NSF grant DMS 0653946
2000 Mathematics Subject Classification: 05A99, 05C35, 05D99

same statement holds with $c = 1/2$, thus confirming an old conjecture of Erdős. They also proved similar results for complete graphs.

In this paper we extend the results of Erdős and Lovász-Simonovits by proving such statements for the broader class of color critical graphs, which are graphs that contain an edge whose removal reduces their chromatic number[1]. In many ways our proof is independent of the specific structure of F.

The main new tool we use is the graph removal lemma, which is an easy consequence of Szemerédi's regularity lemma. In subsequent papers [6, 7] we will extend these results to hypergraphs. The novelty in this project is the use of the removal lemma to count substructures in (hyper)graphs rather precisely.

We often associate a graph with its edge set. Given graphs F, H, where F has f vertices, a copy of F in H is a subset of f vertices and $|F|$ edges of H such that the subgraph formed by this set of vertices and edges is isomorphic to F. In other words, if we denote $Aut(F)$ to be the number of automorphisms of F, then the number of copies of F in H is the number of edge-preserving injections from $V(F)$ to $V(H)$ divided by $Aut(F)$.

Theorem 1. (Graph Removal Lemma) *Let F be a graph with f vertices. Suppose that an n vertex graph H has at most $o(n^f)$ copies of F. Then there is a set of edges in H of size $o(n^2)$ whose removal from H results in a graph with no copies of F.*

Say that a graph F is r-critical if it has chromatic number $r + 1$ and it contains an edge whose deletion reduces the chromatic number to r. As usual, we define the Turán number $\mathrm{ex}(n, F)$ to be the maximum number of edges in an n vertex graph that contains no copy of F as a (not necessarily induced) subgraph. The Turán graph $T_r(n)$ is the n vertex r-partite graph with the maximum number of edges; its parts all have size $\lceil n/r \rceil$ or $\lfloor n/r \rfloor$. Let

$$t_r(n) = |T_r(n)| = \sum_{1 \le i < j \le r} \left\lfloor \frac{n+i-1}{r} \right\rfloor \left\lfloor \frac{n+j-1}{r} \right\rfloor .$$

Since $\chi(T_r(n)) = r$ it does not contain any r-critical graph F, and consequently $\mathrm{ex}(n, F) \ge t_r(n)$. Simonovits [8] proved that if n is sufficiently large, then we have equality. In other words, every n vertex graph $(n > n_0)$ with $t_r(n) + 1$ edges contains at least one copy of F. We extend his result by proving that there are many copies and determining this optimal number. In fact, the number of copies is the number one gets by adding an edge to $T_r(n)$.

[1]Note that a color critical graph is often defined as one with all proper subgraphs having lower chromatic number, but our definition is slightly different

Definition 2. *Fix $r \geq 2$ and let F be an r-critical graph. Then $c(n, F)$ is the minimum number of copies of F in the graph obtained from $T_r(n)$ by adding one edge.*

Note that for any fixed F, computing $c(n, F)$ is just a finite process (see Lemma 5 in the next section). Our result below therefore gives an explicit formula for each color critical F, even though this formula may be very complicated.

Theorem 3. *Fix $r \geq 2$ and an r-critical graph F. There exists $\delta = \delta_F > 0$ such that if n is sufficiently large and $1 \leq q < \delta n$, then every n vertex graph with $t_r(n) + q$ edges contains at least $q\, c(n, F)$ copies of F.*

Theorem 3 is asymptotically sharp, in that for every $q < \delta n$, there exist graphs with $t_r(n)+q$ edges and at most $(1 + O(1/n))q\, c(n, F)$ copies of F. To see this, simply add a matching of size q to the appropriate part of $T_r(n)$. Each new edge lies in precisely $c(n, F)$ copies of F that contain only one new edge, giving a total of $q\, c(n, F)$ copies. If F has f vertices, then the number of copies of F that contain at least two new edges is at most $O(q^2 n^{f-4})$. It is easy to see that $c(n, F) = \Theta(n^{f-2})$ (see Lemma 5 in the next section for more details), and since $q < n$, we obtain $q^2 n^{f-4} = O(1/n)q\, c(n, F)$ as desired.

In many instances Theorem 3 is sharp. Let us examine two special cases.

Odd cycles. Fix $k \geq 1$ and let $F = C_{2k+1}$. Then we quickly see that

$$c(n, F) = \lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 1) \cdots (\lfloor n/2 \rfloor - k + 1)(\lceil n/2 \rceil - 2) \cdots (\lceil n/2 \rceil - k).$$

where we interpret the second product as empty if $k = 1$. Moreover, if we add a star within one of the parts of $T_2(n)$, then it is easy to see that no copy of F contains two edges of the star. Hence Theorem 3 is sharp in the case $F = C_{2k+1}$. Even the simple case of counting C_5's was not previously known.

K_4 minus an edge. Let F be the graph obtained from K_4 by deleting an edge. Then it is easy to see that $c(n, F) = \binom{\lfloor n/2 \rfloor}{2}$ and this is sharp by adding a matching to one part.

For any fixed $\varepsilon > 0$ and $1 \leq q < n^{1-\varepsilon}$, the proof of Theorem 3 actually produces a vertex that lies in $q\, c(n, F)$ copies of F or $(1 - \varepsilon)q$ edges that each lie in $(1 - \varepsilon)c(n, F)$ copies of F. Indeed, if the number of copies of F is $\Omega(n^f)$, then by averaging we immediately obtain a vertex in at least $\Omega(n^{f-1})$ copies of F. This is greater than $qc(n, F)$, since it will be shown later (Lemma 5) that $c(n, F)$ has order of magnitude n^{f-2}. On the other hand, if the number of copies of F is $o(n^f)$, then the proof of Theorem 3 will apply. Such information about the distribution of the copies of F does not seem to follow from the methods of [1, 2, 4], even for cliques.

Throughout the paper, Roman alphabets (e.g. r, s, n) denote integers and Greek alphabets (e.g. $\alpha_F, \beta_F, \gamma_F, \varepsilon, \delta$) denote reals. Given a set of pairs H, let $d_H(v)$ be the number of pairs in H containing v. So if we view H as a graph, then $d_H(v)$ is just the degree of vertex v.

2 Three lemmas

In this section we will prove three technical lemmas needed in the proof of Theorem 3.

Lemma 4. *Suppose that $r \geq 2$ is fixed, n is sufficiently large, $s < n$ and $n_1 + \cdots + n_r = n$. If*

$$\sum_{1 \leq i < j \leq r} n_i n_j \geq t_r(n) - s,$$

then $\lfloor n/r \rfloor - s \leq n_i \leq \lceil n/r \rceil + s$ for all i.

Proof. The result is certainly true for $s = 0$ by definition of $t_r(n)$ and the easy fact that $\sum_{1 \leq i < j \leq r} n_i n_j$ is maximized when $|n_i - n_j| < 1$ for all i. So assume that $s \geq 1$ and let us proceed by induction on s. It is more convenient to prove the contrapositive, so assume that for some i, either $n_i > \lceil n/r \rceil + s$ or $n_i < \lfloor n/r \rfloor - s$ and we wish to prove that $\sum_{1 \leq i < j \leq r} n_i n_j < t_r(n) - s$. We may assume that $n_1 \geq \cdots \geq n_r$ and that $n_1 > \lceil n/r \rceil + s$ (the case $n_r < \lfloor n/r \rfloor - s$ is symmetrical and has an almost identical proof). Define $n'_1 = n_1 - 1, n'_r = n_r + 1$ and $n'_i = n_i$ for $1 < i < r$. Then $\sum n'_i = n$ and we certainly have $n'_i > \lceil n/r \rceil + (s-1)$ for some i. By the induction hypothesis,

$$\sum_{1 \leq i < j \leq r} n'_i n'_j < t_r(n) - (s-1).$$

We also have $n_1 \geq \lceil n/r \rceil + s + 1$ and $n_r \leq \lfloor n/r \rfloor$. Consequently,

$$\begin{aligned} \sum_{1 \leq i < j \leq r} n_i n_j &= \sum_{1 \leq i < j \leq r} n'_i n'_j + \lfloor n/r \rfloor - (\lceil n/r \rceil + s) \\ &\leq \sum_{1 \leq i < j \leq r} n'_i n'_j - s \\ &< t_r(n) - (s-1) - s \leq t_r(n) - s \end{aligned}$$

This completes the proof of the lemma. □

Lemma 5. *Fix $r \geq 2$ and an r-critical graph F with f vertices. There are positive constants α_F, β_F such that if n is sufficiently large, then*

$$|c(n,F) - \alpha_F n^{f-2}| < \beta_F n^{f-3}.$$

In particular, $(\alpha_F/2)n^{f-2} < c(n,F) < 2\alpha_F n^{f-2}$.

Proof. We may assume that $r|n$. Indeed, suppose we could prove the result in the case $r|n$ and we are given r that does not divide n. The graph $T_r(n)$ has parts of size $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$, so we can add at most one vertex to $r-1$ parts so that all parts have size $\lceil n/r \rceil$. Also add edges from the new vertices to all parts distinct from the one that they lie in. The resulting graph is $T_r(n')$ with $n < n' < n+r$ vertices and $r|n'$. So we have

$$c(n,F) \leq c(n',F) \leq c(n+r,F) < \alpha_F(n+r)^{f-2} + \beta_F(n+r)^{f-3} < \alpha_F n^{f-2} + (fr\alpha_F + 2\beta_F)n^{f-3}$$

where the last inequality follows since n is large. The theorem therefore holds with α_F and $\beta'_F = fr\alpha_F + 2\beta_F$. A similar argument gives the required lower bound on $c(n,F)$.

Let us write an explicit formula for $c(n,F)$. Let H be obtained from $T_r(n)$ by adding one edge xy in the first part. Say that an edge $uv \in F$ is good if $\chi(F - uv) = r$. Let χ_{uv} be a proper r-coloring of $F - uv$ such that $\chi_{uv}(u) = \chi_{uv}(v) = 1$. Every proper r-coloring of $F - uv$ gives the same color to u, v, since $\chi(F) > r$. Let x^i_{uv} be the number of vertices of F excluding u, v that receive color i. An edge preserving injection of F to H is obtained by choosing a good edge uv of H, mapping it to xy, then mapping the remaining vertices of F to H such that no two adjacent vertices get mapped to the same part of H. Such a mapping is given by a coloring χ_{uv}, and the number of mappings associated with χ_{uv} is just the number of ways the vertices colored i can get mapped to the ith part of H. The vertices mapped to the first part cannot get mapped to x, y since these have been taken by u, v and there are two ways to map $\{u,v\}$ to $\{x,y\}$. Altogether we obtain

$$c(n,F) = \frac{1}{Aut(F)} \sum_{uv\, good} \sum_{\chi_{uv}} 2(n/r-2)_{x^1_{uv}} \prod_{i=2}^{r} (n/r)_{x^i_{uv}}.$$

Expanding this expression, we get a sum of polynomials in n of degree $x^1_{uv} + \sum_{i=2}^r x^i_{uv} = f-2$ with positive leading coefficients. Consequently, $c(n,F)$ is a polynomial in n of degree $f-2$ with positive leading coefficient. Let α_F be this coefficient. Since n is sufficiently large, and the coefficients of $c(n,F)$ are all fixed independent of n, we can bound the absolute value of the contribution of all the other terms by $\beta_F n^{f-3}$ for some positive β_F that depends only on F. This completes the proof. $\square$

Given integers $n_1, \ldots, n_r$, let $c(n_1, \ldots, n_r, F)$ be the number of copies of F in the graph obtained from the complete r-partite graph with parts $n_1, \ldots, n_r$ by adding an edge to the part of size n_1.

Lemma 6. *Let F be an r-critical graph. There is a positive constant γ_F depending only on F such that the following holds for n sufficiently large. If $n_1 + \cdots + n_r = n$ with $\lfloor n/r \rfloor - s \leq n_i \leq \lceil n/r \rceil + s$ and $s < n/3r$, then*

$$c(n_1, \ldots, n_r, F) \geq c(n, F) - \gamma_F s n^{f-3}.$$

Proof. If $s = 0$, then the result holds by the definition of $c(n, F)$, so assume that $s \geq 1$. Let H be the graph obtained from the complete r-partite graph with parts $n_1 \geq \ldots \geq n_r$ by adding an edge xy to the part of size n_1. Since $|n_1 - n_r| \leq 2s + 1$, we can remove at most $2s + 1 \leq 3s$ vertices (excluding xy) from each part of H so that each part has size n_r. The resulting graph H' satisfies $H' - xy \cong T_r(n')$ with $n' \geq n - 3rs$. Since n is sufficiently large and $s < n/3r$, we have

$$\begin{aligned}
(n - 3rs)^{f-2} &\geq n^{f-2} - \sum_{i=0}^{f-3} \binom{f-2}{i} n^i (3rs)^{f-2-i} \\
&= n^{f-2} - \sum_{i=0}^{f-3} \binom{f-2}{i} (3r)^{f-2-i} s n^i s^{f-3-i} \\
&> n^{f-2} - 3rsn^{f-3} \sum_{i=0}^{f-3} \binom{f-2}{i} \\
&> n^{f-2} - 2^f rsn^{f-3}.
\end{aligned}$$

Put $\gamma_F = \alpha_F 2^f r + 2\beta_F$. Lemma 5 now gives

$$\begin{aligned}
c(n_1, \ldots, n_r, F) &\geq c(n - 3rs, F) \\
&> \alpha_F (n - 3rs)^{f-2} - \beta_F (n - 3rs)^{f-3} \\
&= \alpha_F n^{f-2} + \beta_F n^{f-3} - \gamma_F s n^{f-3} \\
&> c(n, F) - \gamma_F s n^{f-3}
\end{aligned}$$

and the proof is complete. □

3 Proof of Theorem 3

In this section we will prove Theorem 3. We need the following stability result proved by Erdős and Simonovits [8] (the case $r = 3$ with a proof sketch for $r > 3$ also appeared in [3]).

Theorem 7. (Erdős-Simonovits Stability Theorem [8]) *Let $r \geq 2$ and F be a fixed r-critical graph. Let H be a graph with n vertices and $t_r(n) - o(n^2)$ edges that contains no copy of F. Then there is a partition of the vertex set of H into r parts so that the number of edges contained within a part is at most $o(n^2)$. In other words, H can be obtained from $T_r(n)$ by adding and deleting a set of $o(n^2)$ edges.*

Remark. The $o(1)$ notation above should be interpreted in the obvious way, namely $\forall \eta, \exists \xi, n_0$ such that if $n > n_0$ and $|H| > t_r(n) - \xi n^2$, then $H = T_r(n) \pm \eta n^2$ edges. We will not explicitly mention the role of ξ, η when we use the result, but it should be obvious from the context. We will also assume that ξ is much smaller than η, since if the result holds for ξ, then it also holds for $\xi' < \xi$. Similar comments apply for applications of Theorem 1.

Proof of Theorem 3. Our constants $\delta_i, \varepsilon_i, \varepsilon$ will enjoy the following hierarchy:

$$1/n \lll \delta_1 \lll \delta_2 \ll \delta_3 \ll \delta_4 \ll \varepsilon_4 \ll \varepsilon_3 \ll \varepsilon_2 \ll \varepsilon_1 \ll \varepsilon \ll 1.$$

The notation $\xi \lll \eta$ above means that we are applying some theorem with input η and output ξ (the quantification is $\forall \eta, \exists \xi$) and $\xi \ll \eta$ simply means that ξ is a sufficiently small function of η that is needed to satisfy some inequality in the proof. In addition, we require

$$\varepsilon < \frac{\alpha_F}{4(\gamma_F + 2^{f^2})},$$

where α_F comes from Lemma 5 and γ_F comes from Lemma 6. Also, n is sufficiently large that Lemma 4, 5, 6 apply whenever needed. We emphasize that ε_4 is an absolute constant that depends only on F. Set $\delta = \varepsilon_4/4\alpha_F$ and suppose that $1 \leq q < \delta n$. Let H be an n vertex graph with $t_r(n) + q$ edges. Write $\#F$ for the number of copies of F in H.

If $\#F \geq n^{f-1/2}$, then since $c(n, F) < 2\alpha_F n^{f-2}$ and $q < (\varepsilon_4/4\alpha_F)n$, we have

$$\#F \geq n^{f-1/2} > (\varepsilon_4/4\alpha_F)n(2\alpha_F n^{f-2}) > q\, c(n, F)$$

and we are done. So assume that $\#F < n^{f-1/2} = (1/n^{1/2})n^f$. Since n is sufficiently large, by the Removal lemma there is a set of at most $\delta_1 n^2$ edges of H whose removal results in a graph H' with no copies of F. Since $|H'| > t_r(n) - \delta_1 n^2$, by Theorem 7, we conclude that there is an

r-partition of H' (and also of H) such that the number of edges contained entirely within a part is at most $\delta_2 n^2$. Now pick an r-partition $V_1 \cup \ldots \cup V_r$ of H that maximizes $e(V_1, \ldots, V_r)$, the number of edges that intersect two parts. We know that $e(V_1, \ldots, V_r) \geq t_r(n) - \delta_2 n^2$, and an easy calculation also shows that each V_i has size $n_i = n/r \pm \delta_3 n$.

Let B (bad) be the set of edges of H that lie entirely within a part and let G (good) be the set of edges of H that intersect two parts, so $G = H - B$. Let M (missing) be the set of pairs which intersect two parts that are not edges of H. Then $G \cup M$ is r-partite so it has at most $t_r(n)$ pairs and

$$t_r(n) + q - |B| + |M| = |(H - B) \cup M| = |G \cup M| = |G| + |M| \leq t_r(n).$$

Consequently,

$$q + |M| \leq |B| \leq \delta_2 n^2.$$

Also, $|H| = |G| + |B|$ so we may suppose that $|G| = t_r(n) - s$ and $|B| = q + s$ for some $s \geq 0$. For an edge $e \in B$, let $F(e)$ be the number of copies of F in H containing the unique edge e from B.

If $s = 0$, then $G \cong T_r(n)$ and $F(e) \geq c(n, F)$ for every $e \in B$ (by definition of $c(n, F)$) so we immediately obtain $\#F \geq |B|c(n, F) = qc(n, F)$.

We may therefore assume that $s \geq 1$. Partition $B = B_1 \cup B_2$, where

$$B_1 = \{e \in B : F(e) > (1 - \varepsilon)c(n, F)\}.$$

A *potential* copy of F is a copy of F in $G \cup M \cup B = H \cup M$ that uses exactly one edge of B.

Claim 1. $|B_1| \geq (1 - \varepsilon)|B|$

Proof of Claim. Suppose to the contrary that $|B_2| \geq \varepsilon|B|$. Pick $e \in B_2$. Assume wlog that $e \subset V_1$. By Lemma 6 (observing that $n_i = n/r \pm \delta_3 n$ and $\delta_3 n < n/3r$), the number of potential copies of F containing e is

$$c(n_1, \ldots, n_r, F) \geq c(n, F) - \gamma_F(\delta_3 n)n^{f-3} > c(n, F) - \gamma_F \delta_3 n^{f-2} > (1 - \delta_4)c(n, F).$$

At least $(\varepsilon/2)c(n, F)$ of these potential copies of F have a pair from M, for otherwise

$$F(e) > c(n_1, \ldots, n_r, F) - (\varepsilon/2)c(n, F) > (1 - \delta_4 - \varepsilon/2)c(n, F) > (1 - \varepsilon)c(n, F)$$

which contradicts the definition of B_2. First, suppose that for at least $(\varepsilon/4)c(n, F)$ of these potential copies of F, there is a pair from M that does not intersect e. The number of times each such pair from M is counted is at most the number of ways to choose the remaining $f-4$ vertices of the potential copy of F (e is fixed), and then $|F| < f^2$ pairs among the chosen vertices so that the resulting vertices and pairs form a graph isomorphic to F. There are at most $2^{f^2} n^{f-4}$ ways to do this. We obtain the contradiction

$$\frac{\varepsilon\alpha_F}{82^{f^2}} n^2 = \frac{(\varepsilon/8)\alpha_F n^{f-2}}{2^{f^2} n^{f-4}} < \frac{(\varepsilon/4)c(n,F)}{2^{f^2} n^{f-4}} \leq |M| < \delta_2 n^2,$$

where the first strict inequality follows from Lemma 5.

We may therefore assume that for at least $(\varepsilon/4)c(n,F)$ of these potential copies of F, there is a pair from M that intersects e. Each such pair is counted at most $2^{f^2} n^{f-3}$ times, so we conclude that there exists $x \in e$ with

$$d_M(x) \geq \frac{(\varepsilon/8)c(n,F)}{2^{f^2} n^{f-3}} > \frac{(\varepsilon/8)(\alpha_F/2)n^{f-2}}{2^{f^2} n^{f-3}} = \frac{\varepsilon\ \alpha_F}{162^{f^2}} n > \varepsilon_1 n.$$

Let

$$A = \{v \in V(H) : d_M(v) > \varepsilon_1 n\}.$$

We have argued above that every $e \in B_2$ has a vertex in A. Consequently,

$$2\sum_{v\in A} d_{B_2}(v) \geq 2|B_2| \geq 2\varepsilon|B| > 2\varepsilon|M| \geq \varepsilon \sum_{v\in A} d_M(v) > \varepsilon|A|\varepsilon_1 n,$$

and there exists a vertex $u \in A$ such that $d_{B_2}(u) \geq (\varepsilon\varepsilon_1/2)n > \varepsilon_2 n$. Assume wlog that $u \in V_1$.

By the choice of the partition, we may assume that u has at least $\varepsilon_2 n$ neighbors in V_i for each $i > 1$, otherwise moving u to V_i increases $e(V_1, \ldots, V_r)$. For each $i = 1, \ldots, r$, let V_i' be a set of $\lceil \varepsilon_2 n \rceil$ neighbors of u in V_i and for $i = 1$, enlarge V_1' by letting $u \in V_1'$. Let $n_i' = |V_i'|$. Pick $v \in V_1' - \{u\}$. Then the number of copies of F in the complete r-partite graph $K(V_1', \ldots, V_r')$ together with edge $e = uv$ is by definition at least

$$c(r\lceil \varepsilon_2 n \rceil, F) > \alpha_F (r\varepsilon_2 n)^{f-2} - \beta_F (r\lceil \varepsilon_2 n \rceil)^{f-3} > \varepsilon_3 n^{f-2}$$

for suitable ε_3 depending only on F. Let us sum this inequality over all such $e = uv \in B_2$ with $v \in V_1'$. Since $\delta = \varepsilon_4/4\alpha_F$ and $c(n,F) < 2\alpha_F n^{f-2}$, we obtain at least

$$(|V_1'| - 1)\varepsilon_3 n^{f-2} \geq (\varepsilon_2 n)\varepsilon_3 n^{f-2} > \varepsilon_4 n^{f-1} = 4\delta\alpha_F n^{f-1} > 4q\alpha_F n^{f-2} > 2q\, c(n,F)$$

potential copies of F containing u. At least half of these potential copies of F must have a pair from M, otherwise we are done. This pair from M cannot be incident with u, since u is adjacent to all vertices in $K(V_1', \ldots, V_r')$ (other than itself). Hence this pair from M misses u. Each such pair is counted at most $2^{f^2} n^{f-3}$ times, so we obtain the contradiction

$$\frac{\varepsilon_4}{2^{f^2+1}} n^2 = \frac{\varepsilon_4 n^{f-1}}{2^{f^2+1} n^{f-3}} < |M| < \delta_2 n^2.$$

This concludes the proof of the Claim. □

If $s \geq 4\varepsilon q$, then counting copies of F from edges of B_1 and using Claim 1 we get

$$\begin{aligned}
\#F \geq \sum_{e \in B_1} F(e) &\geq \sum_{e \in B_1} (1-\varepsilon) c(n, F) \\
&\geq |B_1|(1-\varepsilon) c(n, F) \\
&\geq (1-\varepsilon)^2 |B| c(n, F) \\
&> (1-2\varepsilon)(q+s) c(n, F) \\
&\geq (q + 2\varepsilon q - 8\varepsilon^2 q) c(n, F) \\
&> q\, c(n, F).
\end{aligned}$$

So we may assume that $s < 4\varepsilon q < q < n/3r$. Recall that $n_i = |V_i|$. Then

$$t_r(n) - s = |G| \leq \sum_{1 \leq i < j \leq r} n_i n_j.$$

Now Lemma 4 implies that for each i,

$$\lfloor n/r \rfloor - s \leq n_i \leq \lceil n/r \rceil + s.$$

Observe that $|M| \leq s$ for otherwise $|G \cup M| > t_r(n)$ which is impossible. Pick $e \in B$ and assume wlog that $e \subset V_1$. The number of potential copies of F containing e is by definition $c(n_1, \ldots, n_r, F)$. Now Lemma 6 implies that

$$c(n_1, \ldots, n_r, F) \geq c(n, F) - \gamma_F s n^{f-3}.$$

Not all of these potential copies of F are in H, in fact, a pair from M lies in at most $2^{f^2} n^{f-3}$ potential copies counted above (we may assume that the pair intersects e otherwise it is counted at most $2^{f^2} n^{f-4}$ times). We conclude that

$$F(e) \geq c(n_1, \ldots, n_r, F) - 2^{f^2} n^{f-3} |M| \geq c(n, F) - \gamma_F s n^{f-3} - 2^{f^2} s n^{f-3}.$$

Since $s < q$ this implies that

$$\begin{aligned}\#F \geq \sum_{e\in B} F(e) &\geq (q+s)(c(n,F) - \gamma_F s n^{f-3} - 2^{f^2} s n^{f-3}) \\ &> q\,c(n,F) + s\,c(n,F) - 2q(\gamma_F s n^{f-3} + 2^{f^2} s n^{f-3}).\end{aligned}$$

As $s < q < \delta n < \varepsilon n$ and $\varepsilon < \alpha_F/(4(\gamma_F + 2^{f^2}))$, we have the bound

$$2q(\gamma_F + 2^{f^2}) s n^{f-3} < 2\varepsilon(\gamma_F + 2^{f^2}) s n^{f-2} < s(\alpha_F/2) n^{f-2} < s\,c(n,F)$$

This shows that $\#F > q\,c(n,F)$ and completes the proof of the theorem. □

4 Acknowledgments

I wish to thank Hemanshu Kaul who was instrumental in starting this project by posing a very general question about the number of copies of certain subgraphs that one is guaranteed to find on a vertex. Initial discussions between Kaul and the author led to a proof of a weaker version of Theorem 3 that found $(1-o(1))q\,c(n,F)$ copies of F. I am also very grateful to a referee for pointing out some minor errors in an earlier version of this paper.

References

[1] P. Erdős, On a theorem of Rademacher-Turán, Illinois Journal of Math, 6, (1962), 122–127

[2] P. Erdős, On the number of complete subgraphs contained in certain graphs, Magy. Tud. Acad. Mat. Kut. Int. Közl. 7, (1962), 459–474 005), no. 4, 467–484.

[3] P. Erdős, Some recent results on extremal problems in graph theory. Results. 1967 Theory of Graphs (Internat. Sympos., Rome, 1966) pp. 117–123.

[4] L. Lovász, M. Simonovits, On the number of complete subgraphs of a graph. II, Studies in pure mathematics, 459–495, Birkhuser, Basel, 1983.

[5] W. Mantel, Problem 28, Winkundige Opgaven 10 (1907), 60-61.

[6] D. Mubayi, Counting substructures II: triple systems, submitted

[7] D. Mubayi, Counting substructures III: quadruple systems, submitted

[8] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, 1968 Theory of Graphs (Proc. Colloq., Tihany, 1966) pp. 279–319 Academic Press, New York